

学校编码: 10384

分类号_____ 密级_____

学号: 10220131152392

UDC_____

廈門大學

硕 士 学 位 论 文

基于语料库技术的副词注释研究

A Study of Adverb Annotation Based on
Corpus Technology

黄 超

指导教师姓名: 郑泽芝 教授

专 业 名 称 : 语言学及应用语言

论文提交日期: 2016 年 3 月

论文答辩时间: 2016 年 5 月

学位授予日期: 2016 年 月

答辩委员会主席:

评阅人:

2016 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

词典是学习者学习一种语言的必备工具。汉语副词是一个半开放的词类，意义较空灵，用法较灵活，对汉语非母语学习者而言有一定的难度，因此，汉语副词注释研究一直是辞书研究的重要一环。本文选取了《现代汉语词典》（第六版）、《商务馆学汉语》（2007 版）、《现代汉语常用词用法词典》（2005 版）三部具有代表性的词典为研究对象，基于功能原则，制订了含有 54 个标记项目的副词注释标记集，对三本词典中的 1925 个副词注释语言，进行了标记加工，形成了副词注释项目数据库，并以此为基础，对不同类型的 3 部词典副词注释实态，进行了系统调查和对比分析。为了考察词典注释项目的针对性，本研究还对 HSK 作文语料库中的 4859 条副词偏误语料进行了分析。

本研究分别从宏观和微观两个维度，对词典副词的收词、注释项目分布、两类词典注释项目设置和注释程度对比，中介语偏误和注释项目的对照等方面进行了研究，在辨明异同的基础上，总结出了内外向型汉语学习词典副词注释的特点及不足，并提出了相关建议。

本文研究的特色有：

一. 以三部词典全部副词注释语言的精加工语料库为研究基础，对副词的注释项目分布实态，进行了系统、全面调查；

二. 以语料库技术提供的量化数据为研究基础，突破了传统的举例论证或抽样调查，增强了研究的可靠性；

三. 以 HSK 动态作文语料库为参照，将外向型汉语学习词典副词注释研究从单角度研究，拓展到了互动的双关照层面，这也是数据驱动研究方法一次实践。

本文发现的问题及获得的结论有：

一. 在收词方面，理解性词典《现代汉语词典》以现实性、典范性为主要的收词原则，不仅以常用性为主要关注点，收词范围广泛，数量繁多。外向型汉语学习词典《商务馆学汉语词典》、《现代汉语常用词用法词典》以常用性为主要收词原则，用户意识浓厚。但三部词典副词收词都存在不尽全面之处。

二. 从三部词典的注释项目来看，理解型词典注释比重集中在常规模块和意义模块，体现了以“释义”为中心的注释原则；外向型学习词典的注释除了常规模块、意义模块外，对用法模块关注也引人注目，这可以从两类词典在注释项目

的设置和不同项目注释程度上明显反映出来,体现了学习词典“释义”和“用法”双中心的注释理念。

三. 三部词典在副词释义方式上都存在“词释义”的方式,这不利于学习者学习。《商务馆学汉语词典》和《现代汉语常用词用法词典》在释义方面存在“闭环性不足”、“循环释义”、“准确性不足”等问题。

四. 外向型汉语学习词典在“易混淆词”、“辨析”等用法注释项目比例不高,在注释比例较低的项目上,如:“儿化”、“构词”、“词语色彩”、“参见”、“常用格式”等方面,存在你注他不注的情况。说明这部分内容可用的资源不足,认识不一致,建议在厘清注释语言现状的基础上,建设针对副词的精加工语料库,以改善此状况。

五. 语法是用法的核心,但本研究数据显示在“语法功能”之类的语法注释项目上,两本学习词典的注释率差异明显,《商务馆》不注或少注,《用法》高注释率,考察现有的语法项目,可以发现存在注释语言太过专业化的倾向,这个现象提醒我们:语法信息要否入词典,如何入词典问题值得关注。

六. 从外向型汉语学习词典的共有副词兼类注释情况来看,《商务馆》和《用法》都借鉴了《现汉》的词性标注,却在兼类词词性注释的取舍方面存在着标准模糊、不统一,甚至具有随意性,并存在漏注、错注的现象,建议兼类词性的取舍应考虑词性的常用性。

七. 副词大部分都由名词、形容词和动词转化而来,极少数还可兼代词、数词、介词、助词,并且有许多常用的单字副词可以做语素,这也说明副词作为外国汉语学习者的难点是有依据的。

八. 外向型汉语学习词典副词的“辨析”、“易混淆词”、“连用”、“条件搭配”、“特定搭配”、“常用格式”等项目注释较弱,应当结合动态中介语语料库,从偏误分析的角度增强副词这些项目的注释。

关键词: 外向型学习词典; 副词; 注释项目

ABSTRACT

Dictionaries are necessary tools for learners to learn a language. Chinese adverbs belong to a half-open word class, with relatively empty meaning and flexible usage, which make difficulties for non-native speakers. Therefore, the study on annotations of Chinese adverbs is always an important part of lexicographical study. This paper selected three representative Chinese dictionaries *The Contemporary Chinese Dictionary* (the Six Edition), *The Commercial Press Learner's Dictionary of Contemporary Chinese* (2006) and *A Dictionary of Chinese Usage* (2007) as objects of study, formulated an adverb annotation tag set which included 54 tagged items based on functional principle, tagged the annotation language of 1925 adverbs in listed three dictionaries, and completed a database of adverb annotative items. Based on the above, a systematic investigation and comparative analysis was conducted on adverb annotations of the different typed dictionaries. To examine the pertinence of the annotative items, this study also included and analyzed 4859 adverb errors from *HSK* compositions corpus.

From two dimensions of macro studies and micro studies, this paper compared and analyzed entry inclusion of adverbs in dictionaries, the distribution of annotative items, the design of annotative items and annotative level of the two typed dictionaries, as well as studying the interlanguage errors and correspondence among annotative items. After identifying the similarities and differences, this paper concluded the peculiarity and deficiency on adverb annotations of domestic-oriented and foreign-oriented Chinese dictionaries, and proposed some relevant suggestions.

The research characteristics of this paper include:

Firstly, this paper presented a systematic and comprehensive exploration on the distribution of adverb annotation items based on a fine-finishing corpus of all adverbial annotation language in the three dictionaries.

Secondly, the study based on the quantized data supplied by corpus technology, which broke through the traditions of example discussion or sample survey, as well as

enhancing the reliability of the study.

Thirdly, the study referred to *HSK* dynamic composition corpus, which broadened the single-perspective research on adverb annotation of foreign-oriented Chinese dictionaries to an interactive and correlated level; it is also a practice of data-driven research method.

The conclusions drawn from this paper include:

Firstly, from the point of entry inclusion, the comprehensive dictionary *The Contemporary Chinese Dictionary* employed reality and normativity as its main entry inclusion principles, seldom considering the popularity of words, and had a broad range as well as big quantity of words; while the foreign-oriented Chinese dictionaries *The Commercial Press Learner's Dictionary of Contemporary Chinese* and *A Dictionary of Chinese Usage* emphasized popularity as their main entry inclusion principle, and cultivated strong user consciousness. However, all the three dictionaries have some defects on entry inclusion.

Secondly, from the comparative perspective on annotative items of three dictionaries, the annotative focus of comprehensive dictionary was on regular module and meaning module, which revealed the interpretation-centered annotation principle; while the annotations in foreign-oriented dictionaries paid attention to not only the regular module and meaning module but also the usage module, which was reflected from the design of annotative items and the annotative level of different items, and highlighted the double-centered (interpretation-centered and usage-centered) annotation principle of learner's dictionaries.

Thirdly, all the three dictionaries applied the annotative method of interpreting a word by its synonyms, which is insufficient for learners to learn languages. The existing annotative problems of the *The Commercial Press Learner's Dictionary of Contemporary Chinese* and *A Dictionary of Chinese Usage* included circulation interpretation, insufficient accuracy, and so on.

Fourthly, for foreign-oriented Chinese learner's dictionaries, the ratio of usage

annotative items such as confusing words items and differentiated items is not high; and some annotative items with relatively low annotative ratio, as some aspects of retroflex suffixation, word-formation, color of words, cross-reference entries, common formats, are not well distributed in every entry. Aiming at the existing problems of lacking available annotation resources and clear agreement on annotative principles, a fine-finishing corpus of adverbs should be created to improve the current adverb annotation condition.

Fifthly, grammar is the core of language usage, but the data of this study showed that the annotative rate of two learner's dictionaries on grammar annotative items differentiated distinctly: the annotation of *The Commercial Press Learner's Dictionary of Contemporary Chinese* seldom or never include grammatical function related information, while *A Dictionary of Chinese Usage* has a high rate of including grammatical function related information in annotations. A trend that over-specialized annotative language appeared in dictionaries was found when observing the current grammatical items, which remind us of the questions that whether it is reasonable to include grammatical information to dictionaries and how to include them.

Sixthly, in terms of annotations of the mutual trans-class adverbs in the two foreign-oriented Chinese learner's dictionaries, both *The Commercial Press Learner's Dictionary of Contemporary Chinese* and *A Dictionary of Chinese Usage* have learned the part-of-speech tagging from *The Contemporary Chinese Dictionary*; however, the two dictionaries has vague, not consistent, even casual standards on whether to accept the part-of-speech tagging of trans-class adverbs, as well as some interpretation omissions and mistakes. Based on this, the popularity of certain part-of-speech should be taken into consideration when annotating trans-class adverbs.

Seventhly, most Chinese adverbs were transferred from nouns, adjectives and verbs, very few of them could be used as pronouns, numerals, prepositions, auxiliary words, and many common single-character adverbs could also be used as morphemes, which provided basis for regarding adverbs as a difficult point for foreign Chinese learners.

Eighthly, the adverb annotations of items as differentiation and analysis, confused words, concurrences, collocations and common formats are weak links in foreign-oriented Chinese learner's dictionaries, and these annotations should be enhanced from the perspective of error analysis with the assistance of dynamic interlanguage corpus.

Key Words: Learners' Dictionary; Adverbs; Annotation Items

第一章 绪论	1
1.1 选题依据	1
1.2 研究意义及研究目标	2
1.3 研究综述	4
1.4 研究基础	12
1.5 研究内容	14
1.6 研究方法	15
1.7 研究思路	16
第二章 三部词典的副词收词研究	18
2.1 副词的界定及提取	18
2.2 三部词典副词的收词研究	21
2.2.1 三本词典共有副词	22
2.2.2 仅两部词典共有副词研究	23
2.2.3 三本词典独有副词研究	28
2.3 词典收词特点总结	33
2.3.1 《现汉》的收词特点	33
2.3.2 《商务馆》收词特点	34
2.3.3 《用法》收词特点	35
第三章 三部词典副词注释项目宏观研究	36
3.1 注释项目的标准及分类	36
3.2 三部词典全部注释项目统计	42
3.3 三部词典共有注释项目分析	43
3.4 仅两部词典共有注释项目分析	60
3.4.1 仅《现汉》和《商务馆》共有注释项目	60
3.4.2 仅《现汉》和《用法》共有注释项目分析	62
3.4.3 仅《商务馆》和《用法》共有注释项目分析	64
3.5 各词典独有注释项目分析	65

3.5.1 仅《用法》独有注释项目分析.....	65
3.5.2 仅《商务馆》独有的注释项目分析.....	69
3.5.3 仅《现汉》独有的注释项目分析.....	69
3.6 本章小结.....	69
第四章 典型副词注释研究.....	74
4.1 词典典型副词的选取	74
4.2 典型副词注释项目对比分析	75
4.3 典型副词共有注释项目微观分析.....	91
4.4 本章小结.....	104
第五章 副词偏误与词典注释研究.....	106
5.1 语料库偏误分析对学习词典编纂的重要作用.....	106
5.2 HSK作文语料库典型副词偏误类型分析.....	107
5.3 典型副词错误类型与词典注释项目对比分析.....	118
5.4 基于偏误分析对外向型汉语学习词典副词注释的建议.....	121
5.5 本章小结.....	123
第六章 结语.....	124
6.1 本文研究成果.....	124
6.2 本文研究的不足.....	125
附录 I 各词典副词词表	126
附录 II 本文中HSK作文语料库副词错误标记集	133
参考文献	135
致谢	139

CONTENTS

1 Introduction.....	1
1.1 Reason of topic selection.....	1
1.2 The target and significance of the research	2
1.3 The literature review.....	4
1.4 The research object	12
1.5 The basic structure of the paper	14
1.6 The research steps	15
1.7 The research steps.....	16
2 The Inspect of Adverbs Word-collection in 3 Chinese Dictionaries	18
2.1 Definition and Extraction adverbs	18
2.2 The Inspect of Adverbs Word-collection in 3 Chinese Dictionaries	21
2.2.1 The comparison of the common words in 3 Chinese Dictionaries	22
2.2.2 The comparison of the common words in 2 Chinese Dictionaries	23
2.2.3 The comparison of the unique words in 3 Chinese Dictionaries	28
2.3 Selection of entries Summarization	33
2.3.1 Characteristic of Selection of entries in <i>Xiandai Hanyu Cidian</i>	33
2.3.2 Characteristic of Selection of entries in <i>Commercial Press</i>	34
2.3.3 Characteristic of Selection of entries in <i>Modern Chinese Common Word Usage Dictionary</i>	35
3 Macro scale research in Adverb annotation items.....	36
3.1 Standard and categories of Annotation items	36
3.2 Overall statistics in three dictionaries	42
3.3 Analyzation of shared annotation items	43
3.4 Analyzation of annotation items shared by two dictionaries	60
3.4.1 <i>Xiandai Hanyu Cidian</i> and <i>Commercial Press</i> shared only annotation items.....	60
3.4.2 <i>Xiandai Hanyu Cidian</i> and <i>Commercial Press</i> shared only annotation items.....	62
3.4.3 <i>Xiandai Hanyu Cidian</i> and <i>Modern Chinese Common Word Usage Dictionary</i> shared only annotation items	64
3.5 Special annotation items listed by single dictionary	65
3.5.1 <i>Modern Chinese Common Word Usage Dictionary</i> -only items	65
3.5.2 <i>Commercial Press</i> -only items	69
3.5.3 <i>Xiandai Hanyu Cidian</i> -only items	69

3.6 Chapter Summarization	69
4 Typical Adverb annotation research.....	74
4.1 Typical Adverb Selection	74
4.2 Typical Adverb annotation items comparison.....	75
4.3 Micro scale research in typical Adverb annotation items for 3 dictionaries	91
4.4 Chapter Summarization	104
5 Adverb based errors and Dictionary annotation research.....	106
5.1 Contribution for corpus based biased analysis to learner oriented dictionary editing	106
5.2 Typical adverb analysis for HSK writing corpus.....	107
5.3 Comparison between typical adverb bias and dictionary annotation items	118
5.4 Advices for adverb annotation in non-native learner based dictionary.....	121
5.5 Chapter Summarization	123
6 Summarization	124
6.1 Conclusions.....	124
6.2 Shortage of this study	125
Appendices I	126
Appendices II	133
References	135
Acknowledgement.....	139

第一章 绪论

1.1 选题依据

(一) 汉语缺乏严格形态变化,造成汉语词类划分困难,汉语副词在这个方面问题尤其严重。吕叔湘(1979)讲过:“汉语副词本来就是个大杂烩”^①。它虚实不定,不管是在意义上还是句法功能上,都难以找到一个统一的标准来界定。就副词和其他词类的关系而言,它也有着特殊的一面:首先,汉语的副词由名词、动词或者形容词演化而来,这造成了副词与其他词类在语法性质、语义特征和功能表达上很相似,很难划分出清晰的界限;其次,随着语言的演变,副词可搭配的词语又超出了原来的搭配范围,如现今常用“很中国”、“很女人”等搭配出现,这即使对汉语母语学习者而言,也不失为一大难点。

(二) 随着我国社会经济快速发展,“汉语热”逐渐风靡全球。外向型学习词典是外国人学习汉语的重要工具书,词典副词注释模式是否合理关系到外国人学习汉语效率的高低。虽然近年来我国外向型学习词典的编纂已取得了一些可喜成就,但并不尽如人意。如周小兵(1997)^②、张淑文(2009)^③等学者指出汉语外向型学习词典存在收词不全面、注释不合理等问题。因此,针对外向型学习词典和理解性词典的编纂现状,将我国现有权威的内外向型词典的副词注释模式进行对比研究,同时兼顾对不同外向型学习词典副词注释进行比较研究,从中总结出各自的优缺点,取长补短,有利于推动学习词典的编纂,提高学习词典副词注释的有效性。

(三) 基于汉语内外向型词典副词的对比研究,通过文献查阅我们发现,关注的人不多。这可能是因为副词作为一个纷繁复杂的系统,少有人对整个副词词类进行系统考察;也有人做过《现代汉语词典》副词的释义研究,但往往是基于副词一个小类,主要针对注释模式的释义方面,缺乏对副词注释模式进行全方面的研究分析,也没有将传统理解性词典和外向型词典副词注释进行对比分析。因此,我们将传统内向型词典和外向型词典副词进行对比研究,一方面是对外向型学习词典研究的补充,有助于完善词典副词注释工作,另一方面,这对语言学本

^①《语法修辞讲话》中国青年出版社,1979.第10页

^②周小兵. 对外汉语学习词典的编写[J]. 辞书研究, 1997(1):82-86.

^③张淑文. 对外汉语学习词典中的语法词目化浅谈——以述补结构为例[J]. 辞书研究, 2009(1):82-91.

体研究也是一个重要补充，是语言学理论在学习词典中的新应用。

（四）现行基于汉语内外向型词典副词注释的研究，通过查阅文献，我们发现，现行研究在研究方法上大多采用“举例+定性”的方法，很少采用“定量+系统性分析”的方法，较多进行外向型词典之间的静态对比，很少有基于中介语语料库来考察词典副词注释的问题，因此本文将采用建立语料库、定量+系统性和基于中介语语料库等方法，突破以往的研究方法，对词典副词注释进行立体地、全方位地研究。

本研究拟基于一本传统内向型理解词典：《现代汉语词典》（第六版）（以下简称《现汉》），两本外向型学习词典：《商务馆学汉语词典》（2007）（以下简称《商务馆》）、《现代汉语常用词用法词典》（2005）（以下简称《用法》），构建两类三本词典副词注释语言语料库，对副词的注释模式进行穷尽式地考察，从量化分析的角度，从注释的宏观和微观方面指出外向型词典中副词注释存在的问题；另结合对外汉语教学中留学生的副词偏误问题，总结出副词作为第二语言习得的重点和难点，发现外向型学习词典对这些重点和难点注释时应当偏重的方向，并提出相关建议。

相关名词解释

外向型学习词典与理解性词典

据黄建华（1985）定义：外向型学习词典指的是服务于母语为非词目语言学习者的词典，理解性词典指的是母语为词目语言学习者的词典。外向型学习词典注重的是词的语境意义，为语言使用者提供生成话语或语篇提供所需的词汇信息和使用方法；理解性词典主要针对词语的概念意义进行解释。

注释与释义

从传统词典研究来看，释义一般指词典词目概念义的解释，虽然后来有学者将释义从单纯的词义解释扩大到对语法和语用的解释，但要把语法和语用信息归入释义中，难免会造成概念混淆。而且，在学习词典中，用法对于没有语感的第二语言学习者来说，与词义的解释几乎位于同等地位。注释这个术语除涵盖传统的释义外，也可以包括用法信息，避开了把释义和用法混为一谈，也将用法信息提升到它本来应该具有的重要地位，因此我们倾向于选择注释这个术语。

1.2 研究意义及研究目标

1. 为副词本体的研究提供新思路

通过对词典中副词注释模式的研究,总结出副词注释模式中的常有项和非常有项,并与本体的研究相结合,使得学习者对副词的语义、语用情况等有了进一步的了解,从而为副词的本体研究提供新的思路。

2. 推动外向型学习词典编纂

外国学习者对汉语学习词典的需求日渐兴旺,对学习型词典的要求也日益提高,这也对学习词典的编纂工作提出了更高的要求。注释是词典的核心内容,其合理性、有用性直接影响词典质量高低。一方面,将外向型学习词典的注释模式与理解型词典的注释模式进行比较,总结出理解型词典中可借鉴并运用于外向型学习词典的部分,有利于促进外向型学习词典的编纂;另一方面,对权威型的外向型学习词典的副词注释模式进行分析,总结出各自的优点与不足,可以为以后外向型汉语词典的进一步完善提供可参考的意见,从而推动外向型学习词典的编纂。

3. 推动理解型词典注释模式的不断完善

外向型学习词典既借鉴了理解型词典的编纂,同时又脱胎于理解型词典,有其自身的特点。借鉴外向型学习词典注释模式,有益于理解型词典的编纂朝着科学化、合理化的方向发展,推动理解型词典注释模式的不断完善。

4. 有助于汉语作为第二语言的副词教学与学习

留学生学习汉语时主要借助汉语词典查阅生词,练习对话。如果词典对副词注释不清晰,势必会影响使用者的掌握程度和学习效率。现代汉语副词是特殊的一类词,它既有词汇意义又有语法意义,可以充当状语、补语等句子成分,有的甚至能独立成句,因此,副词教学一直也是对外汉语教学的难点之一,我们通过HSK动态作文语料库查询发现,留学生经常出现这样的错句:

八点上课。现在九点,保罗就来上课。(应为“才来上课”)

八点上课。他七点半才来了。(应为“就来了”)

如果词典对副词的注释能够条理清晰、精确明白,这将有助于留学生的副词学习,提高他们的学习效率和交际水平。

本文基于一本传统理解性词典和两本外向型汉语学习词典,对副词的注释内

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.